DOCUMENT RESUME

ED 390 919 TM 024 326

AUTHOR Gearhart, Maryl; Herman, Joan L.
TITLE Portfolio Assessment: Whose Work Is It? Issues in the Use of Classroom Assignments for Accountability. Evaluation Comment.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.; National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE 95
CONTRACT R117G10027
NOTE 26p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Accountability; Cooperation; Educational Assessment; Educational Change; Inferences; *Performance Based Assessment; *Portfolio Assessment; Portfolios (Background Materials); State Programs; *Student Evaluation; Teacher Student Relationship; Teaching Methods; *Testing Programs; Test Use; *Validity
IDENTIFIERS *Large Scale Programs; Vermont

ABSTRACT

This article examines one of the challenges to the integration of classroom and large-scale portfolio assessment, a challenge posed by the use of a student's classroom portfolio for large-scale assessment of his or her individual competencies. When work is composed with the support of peers, teachers, and parents, whose work is being judged? The validity of inferences drawn from the assessment can be compromised unless the question can be answered. Experiences with portfolio assessment in Vermont and in another study conducted by the Center for Research on Evaluation, Standards, and Student Testing (California) suggest that the quality of student work reflects not only a student's competence, but also the amount and quality of support received from others. Procedures that highlight a student's contribution to the work must be developed, complicated though this will be. Nevertheless, large-scale portfolio assessment programs appear to carry significant benefits for instructional reform. (Contains 5 tables and 73 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 390 919

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. C. BEER

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

PORTFOLIO ASSESSMENT: WHOSE WORK IS IT? ISSUES IN THE USE OF CLASSROOM ASSIGNMENTS FOR ACCOUNTABILITY

M aryl Gearhart & Joan L. Herman

Center for the Study of Evaluation
National Center for Research on Evaluation, Standards, & Student Testing
University of California, Los Angeles

TM024326

UCLA's Center for the Study of Evaluation &
The National Center for Research on Evaluation, Standards, and Student Testing

EVALUATION COMMENT

Winter 1995

Portfolio Assessment: *Whose Work Is It?*
Issues in the Use of Classroom Assignments for Accountability

Maryl Gearhart and Joan L. Herman[1]

Center for the Study of Evaluation
National Center for Research on Evaluation, Standards, and Student Testing
University of California, Los Angeles

To many engaged in educational reform, portfolio assessment captures a vision of assessment integrated with instruction. Concerned about the equity and validity of large-scale assessment, portfolio advocates argue that students' classroom work and their reflections on that work provide a richer and truer picture of students' competencies than do traditional or other on-demand assessments. Concerned about the impact of testing on teaching, advocates point out that, as displays of the products of instruction, portfolios challenge teachers and students to focus on meaningful outcomes. Furthermore, portfolio assessment practices support the assessment of long-term projects over time, encourage student-initiated revision, and provide a context for presentation, guidance, and critique. Given such an ambitious agenda for assessment, instruction and accountability, it is no surprise that what is meant by "portfolio" or "portfolio assessment" varies markedly in practice and purpose.[2]

New & Recent CRESST Technical Reports & Products

See pages 17-23 for listings of new or recently released reports and products from a variety of CRESST researchers.

Shared by most large-scale assessment projects, however, is a commitment to bridge the worlds of public accountability and classroom practice. The goal is to give students, teachers, and policy makers authentic roles in the assessment of students at all levels of an accountability system and to provide data that are appropriate and useful at each level. The portfolio spans one level of decision making to the next, providing detailed evidence at the class-

TM024670

room level of the process and outcomes of student performance to guide instruction and learning, and then supporting more abridged inferences at the large-scale level about the quality of performance and schooling. Integrated with instruction and targeted on high standards for student performance, the portfolio is the bridge that supports reform of classroom practices on the one side and accountability on the other. The vision is enticing, but will it work? Can classroom work be utilized for large-scale, high-stakes assessment?

In this article, we examine one of the challenges to the integration of classroom and large-scale portfolio assessment, a challenge posed by the use of a student's classroom portfolio for large-scale assessment of his or her individual competencies.[3] When raters working outside the classroom context[4] are asked to make judgments about an individual student based on a portfolio of work composed with the support of peers, teachers, and parents, ***whose work is being judged***? We argue that certain answers to this question could threaten the validity of inferences that can be drawn about individual performance from portfolios constructed in the social complexities of classroom life. Thus the investigation of possible answers to the "whose work" question becomes an essential component to the study of the validity of portfolio assessment.

> ...patterns of relationships among on-demand assessments and portfolio assessments raise questions about the validity of test scores.

Concerns regarding the validity of individual student scores are already emerging in the fledgling technical literature on portfolio assessment. Consider, for example, findings from two CRESST efforts to provide evidence of validity. In both of these studies, patterns of relationships among on-demand assessments and portfolio assessments raise questions about the validity of test scores.

- Koretz and his RAND colleagues have been evaluating Vermont's statewide portfolio assessment program since 1990.[5] The Vermont program targets writing and mathematics at Grades 4 and 8 and includes three components for each subject area: year-long student portfolios, "best pieces" drawn from the portfolios, and state sponsored "uniform" tests which are standardized but not necessarily multiple-choice. Patterns of relationships between the results of portfolio assessment and uniform tests in both subjects were problematic (Koretz, Klein, McCaffrey, & Stecher, 1993). While recognizing that portfolios and standard assessment may well emphasize different aspects of a subject domain, the researchers expected correlations between the two types of assessments within a subject to be stronger than those across subject areas. Instead, they found essentially the same level of correlation within and across subject areas: For example, in writing, writing portfolio scores correlated moderately with the standard measure of writing and with the portfolio and standard measures of mathematics.

- Gearhart and Herman have conducted two technical studies of the ratability of classroom writing portfolios. In an initial study, the researchers found no relationship between scores for writing portfolios and for standard writing assessments: Two-thirds of the students classified as competent based on the portfolio score were not so classified on the basis of the standard assessment. Similarly, there was only a weak relationship between contrasting procedures for portfolio scoring: Half the students classified as compe-

Portfolio Assessment: *Whose Work Is It?*
Issues in the Use of Classroom Assignments for Accountability

tent on the basis of the single portfolio score were not so classified when scores for individual pieces were averaged, though correlations between the two kinds of portfolio scores were moderately high (in the .6 range) (Gearhart, Herman, Baker, & Whittaker, 1992, 1993; Herman, Gearhart, & Baker, 1993). In a subsequent comparative study of two writing rubrics, the researchers found a positive relationship between portfolio scores and standard writing assessments for only one of the rubrics (Gearhart, Novak, & Herman, in press).

Granted, these researchers were hampered in their quest for validation by the paucity of technically sound, performance-based criterion measures to which portfolio scores could be compared. Nevertheless, within the constraints set by the current state of the art in performance assessment, findings like those we have illustrated do raise questions about the validity of portfolio scores as measures of individual performance. What factors may have contributed to these weak relationships between portfolio scores and on-demand assessments? No doubt there are many, and each will require further investigation.

> As summed up by a Vermont teacher after rating portfolios for several days: "Whose work is this anyway?"

Consider just two that focus on measurement design: The portfolio and on-demand assessments may have tapped different domains of performance within a subject area; the on-demand and portfolio tasks may have differed in difficulty.[6] The factor that we consider in this paper arises from the classroom context of portfolio assessment. As summed up by a Vermont teacher after rating portfolios for several days: "Whose work is this anyway?"

We begin our discussion by examining the ways in which the nature of classroom work may undermine the validity of "individual" portfolio scores. We illustrate with CRESST data from both an evaluation of a statewide assessment program and a laboratory study of the scorability of elementary writing portfolios. We conclude with a discussion of the implications of the "whose work" issue for portfolio assessment policy and practice.

Whose Work Is It? An Issue for the Validity of Large-Scale Portfolio Assessment

The "whose work is it?" question arises because individual student portfolios are constructed in a social context. Portfolios contain the products of classroom instruction, and good classroom instruction according to current pedagogical and curriculum reforms involves an engaged community of practitioners in a supportive learning process (Camp, 1993; Duschl & Gitomer, 1991; Wolf, D.P., 1989; Wolf, Bixby, Glenn, & Gardner, 1991; Wolf & Gearhart, 1993a, 1993b). Exemplary instructional practice, in short, *supports* student performance. Central to the National Writing Project, for example, is a core instructional model which features multiple stages—prewriting, precomposing, writing, sharing, revising, editing and evaluation. Each of these stages stands for instructional activities that engage a student with resources and with others—related readings, classroom discussions, field trips, idea webs, small group collaboration, outlining, peer review, review and feedback. The socially contexted character of student writing is seen both as a scaffold for students' writing process and a replication of what "real" writing entails, in that writing is often a very social endeavor. Consider as well what is regarded as exemplary portfolio assessment practice. A "portfolio culture" is viewed as "replacing...the entire envelope of assessment...with extended, iterative processes, agreeing that we are

Portfolio Assessment: *Whose Work Is It?*
Issues in the Use of Classroom Assignments for Accountability

interested in what students produce when they are given access to models, criticism, and the option to revise" (Wolf, D. P., 1993, p. 221). Assessment opportunities are available at multiple classroom moments—in the course of the work that may be added to a portfolio, in the construction of the portfolio, and in a presentation of the portfolio, making collaboration, assessment, and revision continual processes within the classroom.

These visions of an engaged community of learners and reviewers have implications for the validity of classroom portfolios for large-scale assessment purposes: The more developed the community, the more engaged others will be in the work tagged with an individual student's name. While the locus of authorship may shift outward from the individual student to the community of writers, the shift is unlikely to be systematic: Others' contributions to students' work are likely to vary across assignments, students, and classrooms. An irony emerges that when the student's work *is* more her own, that work may index practices and curriculum that *lack* certain key features of current reforms.

How is a rater unfamiliar with a student or the classroom context to assign an individual student a score for a portfolio collection that includes assisted or collaborative work? Research by Webb (1993) suggests that an individual's performance in the context of group activity may or may not represent his or her capability. Her finding, for example, that low-ability students had higher scores on the basis of group work than on individual work suggests that a rater's score for a portfolio may overestimate student performance because it constitutes a rating of efforts that were assisted. Alternatively, the rater who is aware that work is assisted may adjust downward the individual's score, again biasing the rating.[7]

Whose Work Is It? Data From CRESST Studies

While questions regarding the roles of authorship and assisted performance in large-scale portfolio assessment have been raised (Condon & Hamp-Lyons, 1991; Gitomer, personal communication, September, 1994; Herman, et al., 1993; Koretz, McCaffrey, Klein, Bell, & Stecher, 1993; Koretz, Stecher, & Deibert, 1992; Koretz, personal communication, September, 1994; Stecher & Hamilton, 1994), they have been neither directly investigated nor widely discussed. As we discuss next, however, preliminary results from the CRESST Vermont studies (Koretz, Stecher, Klein, & McCaffrey, in press) and the laboratory-based studies of portfolio ratability (Gearhart et al., 1992; Gearhart, Herman, Novak, Wolf, & Abedi, 1994; Herman et al., 1993) add some empirical basis for concern. These studies suggest substantial variability in instructional support for students' work, variability which may well compromise the meaning and comparability of scores within as well as between classrooms and schools.

> These studies suggest substantial variability in instructional support for students' work...

Vermont

While the RAND evaluation addresses three broad issues—the actual implementation of the program in schools and classrooms, the program's diverse effects, and the quality of the information yielded by the assessment—of interest here are results from a survey distributed to all fourth- and eighth-grade math teachers during the second year (1992-93) of Vermont's statewide implementation.[8] Results are based on the responses of approximately 52% of the mathematics teachers at Grade 4 ($N = 382$) and 41% at Grade 8 ($N = 137$) (p. 6).

Portfolio Assessment: *Whose Work Is It?*
Issues in the Use of Classroom Assignments for Accountability

Teachers' responses to a number of questions indicated substantial variation in how mathematics portfolios were implemented across classrooms, and consequently substantial variation in how much help and support students received in putting their "best face forward" for the portfolio assessment. Teachers' reported policies on revising best pieces are a first case in point: Although more teachers *encouraged* revision of most best pieces (57%), many teachers departed from this pattern by either *requiring* revision (19%), simply *permitting* it (19%), or generally *prohibiting* it (5%). Similarly, the amount of time students spent revising varied widely. The average time in revision was 30-40 minutes, but in 17% of classrooms students did not revise at all, and in another 15% of classrooms students took more than one full class period to revise a best piece. Provision of time and support for revision clearly represents an aid to performance, and thus students who are not encouraged to revise their best pieces may well be at a disadvantage relative to those students who are provided greater opportunities to revise.

There also was considerable variation in teachers' policies regarding who was permitted to assist students in revising their best pieces (Table 1). One in four teachers did not report assisting their own students in revisions, and a similar proportion did not report permitting students to help each other. Seventy percent of fourth-grade teachers and 39% of eighth-grade teachers forbade parental or other outside assistance. Further complicating these find-

Table 1
Assistance Allowed by Teachers on Best Pieces
(Percentage of Teachers)

		Assistance allowed on which best pieces?				Rules differ for individual students
Source	Grade	None	Some	Most	All	
Teacher	4	27	23	14	16	21
	8	27	32	9	13	19
Other Students	4	34	31	11	12	1
	8	23	39	11	12	15
Parents or others outside school	4[a]	71	13	4	4	8
	8	39	28	8	13	11

[a] Grade-level difference significant at the 5% level ($p<.05$).

ings regarding classroom variation, roughly 10% of teachers reported that their policies regarding assistance varied for different students *within* their classrooms. Teachers' policies also differed with respect to acknowledgment of outside help. Only about 20% of teachers required their students to acknowledge or describe the assistance they received, and, therefore, the raters of most students' portfolios would not know who contributed to the entries or the nature of their assistance.

Finally, the Vermont teachers reported substantially different degrees of influence on students' choices of "best pieces" for their portfolios (Table 2): Some teachers reported playing an equal role with their students in making portfolio selections, while others reported no role at all. Certainly the type and quality of the work that becomes part of a student's portfolio can be influenced by who selects the pieces for inclusion. In particular, since teachers presumably have a better understanding of the scoring criteria than do students, the portfolios of students whose choices were assisted by their teachers may be more likely to show students' capabilities.

Thus the RAND/CRESST study found sizable variations among classrooms in factors such as the amount of revision that was permitted and the extent to which teachers limited assistance from others. These implementation findings may help to explain the weak patterns of relationships between portfolio scores and on-demand assessments, relationships decribed earlier. If some teachers provide (directly or through other adults or students) more help than others, comparisons among the portfolio scores of their students would be clouded by the contributions that others make to a given student's portfolio. Because such factors enter only into portfolio scores and not into scores on a standardized, on-demand assessment, they would tend to weaken the relationships between portfolio scores and scores from on-demand assessments.

Table 2
Who Selects Best Pieces? (Percentage of Teachers)

Who selects best pieces?	Grade 4[a]	Grade 8
Students on their own	21	30
Students with limited teacher input	55	57
Students and teachers have equal role	18	8
Teacher with limited student input	5	3
Teacher	1	1

[a] Grade-level difference significant at the 5% level ($p<.05$).

Portfolio Assessment: *Whose Work Is It?*
Issues in the Use of Classroom Assignments for Accountability

CRESST Laboratory Studies of the Scorability of Writing Portfolios

To document the contributions of others to the writing contained within students' writing portfolios, Gearhart et al. (1993)[9] asked teachers to rate the level of their instructional support for their writing assignments. Data were collected in the spring of 1991 from nine teachers spanning Grades 1 to 6. Each teacher was asked to designate two students at each of three levels of writing competency (high, medium, and low), to collect complete portfolios of all of their work, and for *each* writing assignment to document the instructional support provided during the composing and editing phases.

Ratings were keyed to the same dimensions used at that time to assess students' writing progress (Baker, Gearhart, & Herman, 1991): *Content/Organization* (topic/subtopics or theme, and their structure and development); *Style* (elements of text like descriptive language, word choice, sentence choice, tone, mood, voice, and audience); and *Mechanics* (spelling, grammar, punctuation, and other conventions). The scale points were defined along a continuum from 0 (no support) to 3 (teacher has specified the requirement in detail). Teachers also were asked to rate each assignment in terms of *Copied work* (the extent to which the student's work appeared to be copied from peers or from direct modeling by a teacher or parent) and to estimate the time the child spent on the assignment in hours or fractional parts of hours. The dataset consisted of spring 1991 ratings of 228 assignments from a total of 54 students. The number of assignments per student ranged from 1 to 21, with a modal number of 3. (One teacher returned 14-21 assignments per target student, compared with 1-5 for the remaining eight teachers.[10]) Across all assignments, teachers reported providing generally low to moderate levels of support to their target students, but their reported support differed substantially among students' competency levels: Teachers were far more likely to report providing higher levels of support to their "low" students than to their more able students (Table 3), a finding that raises concerns about the differential meaning of scores that may be assigned to students' portfolios.

The patterns of teachers' reported support differed across the three writing dimensions, reflecting,

Table 3
Percentage of Teachers Reporting Greater Support by Writing Dimension and Student Ability Level

	Student ability level	
Writing dimension	High	Low
Content/organization	34	72
Style	13	55
Mechanics	26	60

Note. A "greater" level of support was defined as ratings of 2 or 3, where 2 indicated *some* guidelines and feedback, and 3 represented *detailed* guidelines and feedback.

it seems, variations in curriculum. In Table 4, we see that teachers' experience with portfolio assessment was related to their patterns of instructional support. The three teachers who had been using portfolios in their classrooms for over a year tended to report providing higher levels of support than did the six teachers who had just begun experimenting with portfolio assessment, and we believe that the more experienced teachers' engagement with a writing process approach contributed to their greater involvement with students' assignments (and/or to their greater perceptions of involvement). Table 5 hints at the ways that teachers' reported levels of support may be related to grade level as well as portfolio experience. While these data are purely an illustration from a very small dataset, we see here two second-grade teachers providing quite different levels of assistance with style vs. mechanics, and two fifth-grade teachers differing more in levels of support for style. Furthermore, the second- and fifth-grade teachers with a year of portfolio experience reported emphases on different writing dimensions—the second-grade teacher more concerned with mechanics, the fifth-grade teacher more concerned with style.

Thus teachers in the Gearhart et al. (1993) study reported variations in instructional practices that were likely to have impacted differentially the quality of student work in the portfolios. As in Vermont, these implementation findings may help to explain the weak relationships between students' portfolio scores and their standard writing assessments.

Reflections and Recommendations

Teacher self-report data from two CRESST studies have produced evidence of variation in how portfolio work is produced and supported. We acknowledge the flaws of the preliminary self-report data that we have presented and fully recognize that further research is needed—studies that employ larger sample sizes and multiple methodologies to verify the variety of support provided to students and the impact of such support on assessed performance. But if findings like these can be substantiated in more systematic research, they suggest that the quality of student work reflects not only a student's

Table 4
Comparison of Teachers With Little vs. One Year Experience With Portfolios: Percentage of Assignments Given Greater Support by Writing Process Dimension

	Dimension		
Portfolio experience	Focus/organization	Style	Mechanics
Little (*n*=6)	54	41	36
One year (*n*=3)	92	82	74

Note. A "greater" level of support was defined as ratings of 2 or 3, where 2 indicated *some* guidelines and feedback, and 3 represented *detailed* guidelines and feedback.

Table 5
Illustrative Comparison of Selected Teachers: Percentage of Assignments Given Greater Support by Writing Process Dimension

Teacher's grade level & portfolio experience	Dimension		
	Focus/organization	Style	Mechanics
Second grade			
Little experience with portfolios	83	58	67
One year experience with portfolios	65	18	100
Fifth grade			
Little experience with portfolios	50	29	21
One year experience with portfolios	72	72	44

Note. A "greater" level of support was defined as ratings of 2 or 3, where 2 indicated *some* guidelines and feedback, and 3 represented *detailed* guidelines and feedback.

competence but also the amount and quality of support received from others. Thus, whose work is the classroom work contained in a student's portfolio? From the preliminary evidence presented here, it seems it may depend—on students' competence and a range of variable circumstances: teachers' methods of instruction, the nature of their assignments, peer and other resources available in the classroom, and home support.

What meaning, then, can a large-scale portfolio assessment program ascribe to student work contained in portfolio collections? A professional—whether a writer, scientist, or educational researcher—who is accustomed to others' input may respond to this question with philosophical reflection or an identity crisis. Indeed, whose work is this article, for example? In what ways does it reflect the writing and research competencies of either of its authors? We value our own opportunities for collaborative work as much as we value the efforts to engage students in authentic communities in the classroom. But, from a measurement perspective, the validity of inferences about student competence based solely on portfolio work appears suspect. While this is not a grave concern for classroom assessment where teachers and students can judge performances with knowledge of their context, the problem is troubling indeed for large-scale assessment purposes where comparability of data is an issue. Under what circumstances, then, can portfolio assessments be used to rank or make serious decisions about students, teachers, schools, or districts?

Portfolio Assessment: *Whose Work Is It?*
Issues in the Use of Classroom Assignments for Accountability

The question requires attention to (a) the purposes of portfolio assessment, (b) the integrated design of portfolio contents, rubric contents, rating procedures, and uses of the results, and (c) a recognition of possible conflicts between the measurement and instructional aims of portfolio assessment. The apparent inverse relationship between support and students' ability level in the Gearhart et al. (1993) study is a telling example in this regard. Certainly if low-performing students are to achieve high standards, it is likely they will need an enriched instructional process to give them the capability for transferable performance—ample models, coaching and mentoring, and multiple opportunities for practice, feedback and revision. But if the work that emerges from this same instructional process is used to assess students' ***individual*** performance, then there will be problems of comparability of scores across students. Can we bridge this apparent gulf between what is required to serve the purposes of classroom instruction and large-scale accountability?

While no easy solutions come to mind, it does appear that any valid assignment of an individual student score to a portfolio for large-scale purposes will require procedures to highlight the student's contribution to the work. Adjustments in either composition of the portfolio or rating procedures, or both, will be necessary to assure comparability of student results. As we outline below, a number of strategies have been suggested, although most have so far been rejected for reasons of feasibility, cost, or violation of certain fundamental portfolio program assumptions, such as instructional freedom, seamless integration of portfolios with instruction, and commitment to honor diversity (D. Gitomer, personal communication, September, 1994; D. Koretz, personal communication, September, 1994; K. Sheingold, personal communication, September, 1994).

- Restrictions could be imposed on work students produce for their portfolios, controlling who is permitted to provide assistance and under what circumstances. These procedures, largely rejected as violations of instructional freedom, would require verification that the controls on assistance were in place.

- Portfolios could be "seeded" with students' responses to a standard performance-based writing assessment; ratings of these entries might be used to adjust overall portfolio scores, or to raise "red flags" when scores for standard assessments are discrepant with other portfolio material. But this option would bring additional complications.

> Portfolio procedures could incorporate strategies for documenting others' assistance and input...

First, many performance-based assessments of writing and reading currently incorporate components of a process approach—for example, shared readings and class discussion, or even peer response—and thus the "seeded" assessments might also need checks that the assistance provided was comparable across students. Second, procedures for adjusting portfolio scores would require consensus on a framework for justifying those procedures: On what grounds can a student's individual writing be compared against his writing supported by others' responses and guidance?

Portfolio procedures could incorporate strategies for documenting others' assistance and input:

Portfolio Assessment: *Whose Work Is It?*
Issues in the Use of Classroom Assignments for Accountability

- Portfolio entries could be accompanied with context descriptions that document the resources that were available to a student and the contributions of others to the process of composing the work. However, this option would require procedures for (a) producing those descriptions (training teachers and/or students to provide comparable information across assignments, students, and classrooms), and (b) using the context information when rating the portfolio material (e.g., will raters first rate the work on its own merits, and then adjust the score after examining evidence of support? Will raters rate the work and let someone else rate the support and make the adjustment? Will raters look at all sources of evidence and make some kind of integrated, summary judgment of the student's individual contribution and competence?).

- Raters could score a sample of portfolios from a given classroom at virtually the same time, to enable them to see how an individual's work compares with or duplicates the responses of others from the class. To date, this procedure has simply been used as an exercise, but it is of interest that the exercise has yielded two different conclusions. In the Vermont project, researchers reviewing mathematics portfolios from one class noticed identically worded key phrases in many students' responses to the same assignment. This suggested that the work had been structured so heavily by the teacher that the responses did not really represent the performance of individual students (S. Klein and D. Koretz, personal communication, September, 1994). In another project, six target students were interviewed in each of several portfolio classrooms, and, in some classrooms, worrisome commonalties in the content of students' writing reflected impressive commonalties in students' *understandings* of the work based on their interview responses; students had learned a great deal from an assignment that emerged from intensive classroom collaboration. How could we tell the difference, then, between common understandings and copying?

There is interest in incorporating assessment of group process.

- Portfolios could be constructed to provide evidence for the very interactive processes that endanger the validity of individual scores assigned to the final product. That is, if students are using resources, and soliciting and making use of input from others, then it makes sense to document and assess students' competencies with these ways of working within a writing community (cf. a current analytic review on methods for group assessment by Webb, 1994).

Similarly, students' unique contributions to group products could be documented:

- The design of portfolio assessment could document more explicitly an individual student's role in a given product. The inclusion of student self-assessments, peer assessments (when the work was collaborative), and teacher assessments could help to clarify a student's unique contribution; a follow-up individual assignment could demonstrate what a student had learned from a collaborative project or a project heavily guided by the teacher. However, once again, little is known about the ways that raters would utilize these inclusions in making a portfolio judgment.

Portfolio Assessment: *Whose Work Is It?*
Issues in the Use of Classroom Assignments for Accountability

Listed here to stimulate thinking about possible solutions, our collated suggestions provide imperfect and somewhat unwieldy answers. The alternative? A large-scale, high-stakes portfolio program could produce individual student scores but do nothing to address the "whose work" problem, thereby ensuring invalid comparisons among students.

This is not to say that large-scale portfolio assessment is without value. Our own research and that of others indicates that large-scale portfolio assessment programs carry significant benefits for instructional reform (Gearhart & Wolf, 1994; Herman & Winters, 1994; Koretz, Stecher, Klein, McCaffrey, & Deibert, 1993; Koretz et al., in press; Sheingold, Heller, & Paulukonis, in press). In Vermont, for example, principals and teachers reported that the portfolio program has induced sizable and diverse changes in instruction that are largely consistent with the goals of reform efforts both in Vermont and nationwide. This is an important consequence even if portfolios fail to provide valid individual measurement. But *can* portfolio assessment provide us with valid indices of student competencies usable for large-scale accountability? There are some promising possibilities: For example, the question of "whose work" may be less troublesome if portfolio assessment results are aggregated for decision making at the school or district level, using matrix sampling and excluding individual-level data. As Moss (1992, 1994) argues, multilevel designs could assign responsibility for individual-level decisions to the local level, where portfolio-based decisions about the capability of individual students can be informed by professional judgment and knowledge of the local school context.

> ...in the hands of skilled teachers, parents, and students, portfolios should provide critical contexts for discussing, assessing and improving the process and the outcomes of students' learning.

Students' portfolio work could also provide invaluable evidence of students' "opportunities to learn." If some controls over inequitable help from sources outside the classroom were in place, then portfolio assessment (at any level of the system) could provide a window on the quality of curriculum and instruction by showing what work students are asked to do and how well they are able to do it. From this perspective, we would expect students' portfolios to show the best of what students can do with help from an effective instructional process.

Finally, in moving forward on large-scale portfolio assessment, we will need to remember that the very complexity of portfolio assessment is at once its strength and its weakness. While portfolios may resist attempts to reduce their contents to simple, reliable individual scores, in the hands of skilled teachers, parents, and students, portfolios should provide critical contexts for discussing, assessing and improving the process and the outcomes of students' learning.

References

Baker, E. L., Gearhart, M., & Herman, J. L. (1991). *The Apple Classrooms of Tomorrow: 1990 evaluation study. Report to Apple Computer, Inc.* Los Angeles: University of California, Center for the Study of Evaluation.

Baker, E. L., Gearhart, M., Herman, J. L., Tierney, R., & Whittaker, A. K. (1991). Stevens Creek portfolio project: Writing assessment in the technology classroom. *Portfolio News, 2*(3), 7-9.

Belanoff, P., & Dickson, M. (Eds.). (1991). *Portfolios: Process and product.* Portsmouth, NH: Heinemann.

Portfolio Assessment: *Whose Work Is It?*
Issues in the Use of Classroom Assignments for Accountability

Calfee, R.C., & Perfumo, P. A. (1992). *A survey of portfolio practices.* Berkeley: University of California, Center for the Study of Writing.

Calfee, R. C., & Perfumo, P. A. (1993a). *Student portfolios and teacher logs: Blueprint for a revolution in assessment* (CSW Tech. Rep. TR 65). Berkeley: University of California, Center for the Study of Writing.

Calfee, R. C., & Perfumo, P. A. (1993b). Student portfolios: Opportunities for a revolution in assessment. *Journal of Reading, 36*, 532-537.

Camp, R. (1990). Thinking together about portfolios. *The Quarterly of the National Writing Project and the Center for the Study of Writing, 12*(2), 8-14, 27.

Camp, R. (1993). The place of portfolios in our changing views of writing assessment. In R. E. Bennett & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment* (pp. 183-212). Hillsdale, NJ: Erlbaum.

Camp, R., & Levine, D. S. (1991). Portfolios evolving: Background and variations in sixth through twelfth grade classrooms. In P. Belanoff & M. Dickson (Eds.), *Portfolios: Process and product* (pp. 194-205). Portsmouth, NH: Heinemann.

Condon, W., & Hamp-Lyons, L. (1991). Introducing a portfolio-based writing assessment: Progress through problems. In P. Belanoff & M. Dickson (Eds.), *Portfolios: Process and product* (pp. 231-247). Portsmouth, NH: Heinemann.

Duschl, R. A., & Gitomer, D. H. (1991). Epistemological perspectives on conceptual change: Implication for educational practice. *Journal of Research in Science Teaching, 28*(9), 839-858.

Freedman, S. (1993). Linking large-scale testing and classroom portfolio assessments of student writing. *Educational Assessment, 1*(1), 27-52.

Gearhart, M., Herman, J. L., Baker, E. L., & Whittaker, A. K. (1992). *Writing portfolios at the elementary level: A study of methods for writing assessment* (CSE Tech. Rep. 337). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Gearhart, M., Herman, J. L., Baker, E. L., & Whittaker, A. (1993). *Whose work is it? A question for the validity of large-scale portfolio assessment* (CSE Tech. Rep. 363). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Gearhart, M., Herman, J. L., Novak, J. R., Wolf, S. A., & Abedi, J. (1994). *Toward the instructional utility of large-scale writing assessment: Validation of a new narrative rubric* (CSE Tech. Rep. 389). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Gearhart, M., Novak, J. R., & Herman, J. L. (in press). *Issues in portfolio assessment: The scorability of narrative collections* (CSE Tech. Rep.). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Gearhart, M., & Wolf, S. A. (1994). Engaging teachers in assessment of their students' writing: The role of subject matter knowledge. *Assessing Writing, 1*(1), 67-90.

Gearhart, M., & Wolf, S. A. (in press). *The student's role in large-scale portfolio assessment: Providing evidence of writing competency. Part I: The purposes and processes of writing* (CSE Tech. Rep.). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Gearhart, M., Wolf, S. A., Burkey, B., & Whittaker A. K. (1994). *Engaging teachers in assessment of their students' narrative writing: Impact on teachers' knowledge and practice* (CSE Tech. Rep. 377). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Gentile, C. (1992). *Exploring new methods for collecting students' school-based writing: NAEP's 1990 portfolio study.* Washington, DC: National Center for Education Statistics.

Gill, K. (Ed.). (1993). *Process and portfolios in writing instruction. Classroom practices in teaching English, Vol. 26.* Urbana, IL: National Council of Teachers of English.

Gitomer, D. H. (1993). Performance assessment and educational measurement. In R. E. Bennett & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement* (pp. 241-263). Hillsdale, NJ: Lawrence Erlbaum.

Glazer, S. M., Brown, C. S., Fantauzzo, P. D., Nugent, D. W., & Searfoss, L. A. (1993). *Portfolios and beyond: Collaborative assessment in reading and writing.* Norwood, MA: Christopher-Gordon.

BEST COPY AVAILABLE

Portfolio Assessment: *Whose Work Is It?*
Issues in the Use of Classroom Assignments for Accountability

Hamp-Lyons, L., & Condon, W. (1993). Questioning assumptions about portfolio-based assessment. *College Composition and Communication, 44*, 176-190.

Herman, J. L., Aschbacher, P.R., & Winters, L. (1992). *A practical guide to alternative assessment.* Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).

Herman, J. L., Gearhart, M., & Aschbacher, P. R. (in press). Portfolios for classroom assessment: Design and implementation issues. In R. C. Calfee (Ed.), *Portfolio assessment.* Hillsdale, NJ: Erlbaum.

Herman, J. L., Gearhart, M., & Baker, E. L. (1993). Assessing writing portfolios: Issues in the validity and meaning of scores. *Educational Assessment, 1*(3), 201-224.

Herman, J. L., & Winters, L. (1994). Portfolio research: A slim collection. *Educational Leadership, 52*(2), 48-55.

Hewitt, G. (1991). *Leading and learning: A portfolio of change in Vermont schools* (Tech. Rep.). Vermont: Governor's Institutes.

Hewitt, G. (1993, May-June). Vermont's portfolio-based writing assessment program: A brief history. *Teachers and Writers, 24*(5), 1-6.

Hiebert, F., & Calfee, R. C. (1992). Assessment of literacy: From standardized tests to performances and portfolios. In A. E. Fastrup & S. J. Samuels (Eds.), *What research says about reading instruction* (pp. 70-100). Newark, DE: International Reading Association.

Hill, R. (1992, June). *Assessments developed in support of educational reform in contrast to assessments developed in support of educational refinement.* Paper presented at the Assessment Conference of the Education Commission of the States, Boulder, CO.

Howard, K. (1990). Making the writing portfolio real. *The Quarterly of the National Writing Project and the Center for the Study of Writing, 12*(2), 4-7, 27.

Koretz, D. (1993). New report of Vermont Portfolio Project documents challenges. *National Council on Measurement in Education Quarterly Newsletter, 1*(4), 1-2.

Koretz, D., Klein, S., McCaffrey, D., & Stecher, B. (1993). *Interim report: The reliability of the Vermont portfolio scores in the 1992 school year* (CSE Tech. Rep. 370). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Koretz, D., McCaffrey, D., Klein, S., Bell, R., & Stecher, B. (1993). *The reliability of scores from the 1992 Vermont Portfolio Assessment Program* (CSE Tech. Rep. 355). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Koretz, D., Stecher, B., & Deibert, E. (1992). *The Vermont Assessment Program: Interim report on implementation and impact, 1991-92 school year* (CSE Tech. Rep. 350). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). *The evolution of a portfolio program: The impact and quality of the Vermont program in its second year (1992-93)* (CSE Tech. Rep. 385). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (in press). The Vermont Portfolio Assessment Program: Findings and implications. *Educational Measurement: Issues and Practice, 13*(3).

Koretz, D., Stecher, B., Klein, S., McCaffrey, D., & Deibert, E. (1993). *Can portfolios assess student performance and influence instruction? The 1991-92 Vermont experience* (CSE Tech. Rep. 371). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

LeMahieu, P. G., Eresh, J. T. & Wallace, R.C. (1992). Using student portfolios for a public accounting. *School Administrator, 49*(11), 8-15.

LeMahieu, P. G., Gitomer, D. H., & Eresh, J. T. (in press). Portfolios in large-scale assessment: Difficult but not impossible. *Educational Measurement: Issues and Practice.*

Meyer, C. A., Schuman, S., & Angello, N. (September, 1990). *Aggregating portfolio data* (White Paper). Lake Oswego, OR: Northwest Evaluation Association.

Mills, R. P. (1989, December). Portfolios capture a rich array of student performance. *The School Administrator*, pp. 8-11.

Mills, R. P., & Brewer, W. R. (1988). *Working together to show results: An approach to school accountability in Vermont.* Monpelier: Vermont Department of Education, October 18/November 10.

Portfolio Assessment: *Whose Work Is It?*
Issues in the Use of Classroom Assignments for Accountability

Moss, P. A. (1992). Portfolios, accountability, and an interpretive approach to validity. *Educational Measurement: Issues and Practice, 11*(3), 12-21.

Moss, P. A. (1994, March). Can there be validity without reliability? *Educational Researcher, 23*(2), 5-12.

Moss, P. A., Beck, J. S., Ebbs, C., Herter, R., Matson, B., Muchmore, J., Steele, D., & Taylor, C. (1991, April). *Further enlarging the assessment dialogue: Using portfolio to communicate beyond the classroom.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Murphy, S., & Smith, M. (1990, Spring). Talking about portfolios. *The Quarterly of the National Writing Project and the Center for the Study of Writing, 12*(2), 1-3, 24-27.

Murphy, S., & Smith, M. (1992). Looking into portfolios. In K. B. Yancey (Ed.), *Portfolios in the writing classroom: An introduction* (pp. 49-61). Urbana, IL: National Council of Teachers of English.

O'Neil, J. (1992). Putting performance assessment to the test. *Educational Leadership, 49*(8), 14-19.

Paulson, F. L., Paulson, P. R., & Meyer, C. A. (1991, February). What makes a portfolio a portfolio? *Educational Leadership, 48*(5), 60-63.

Reidy, E. (1992, June). *What does a realistic state assessment program look like in the near term?* Paper presented at the annual ECS/CDE meeting, Boulder, CO.

Saylor, K., & Overton, J. (1993, March). *Kentucky writing and math portfolios.* Paper presented at the National Conference on Creating the Quality School.

Sheingold, K. (1994, September). *Designing a large-scale portfolio assessment system for classroom consequences.* Presentation at the 1994 CRESST Conference: Getting Assessment Right!, Los Angeles.

Sheingold, K., Heller, J., & Paulukonis, S. (in press). *Actively seeking evidence: Shifts in teachers' thinking and practice through assessment development* (Tech. Rep.). Princeton, NJ: Educational Testing Service, Center for Performance Assessment.

Simmons, J. (1990). Portfolios as large scale assessment. *Language Arts, 67*, 262-268.

Smith, M. A., & Murphy, S. A. (1992, Winter). "Could you please come and do portfolio assessment for us?" *Quarterly of the National Writing Project and the Center for the Study of Writing and Literacy, 14*(1), 14-17.

Stecher, B. M., & Hamilton, E. G. (1994, April). *Portfolio assessment in Vermont, 1992-93: The teachers' perspective on implementation and impact.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Stroble, E. J. (1993, December). Kentucky student portfolios: Expectations of success. *Equity and Excellence in Education, 26*(3), 54-59.

Tierney, R. J. (1992). Portfolios: Windows on learning. *Learning, 2*, 61-64.

Tierney, R. J., Carter, M. A., & Desai, L. E. (1991). *Portfolio assessment in the reading writing classroom.* Norwood, MA: Christopher Gordon.

Valencia, S. W., & Calfee, R. C. (1991). The development and use of literacy portfolios for students, classes, and teachers. *Applied Educational Measurement, 4*, 333-345.

Vermont Department of Education. (1990, September). *Vermont Writing Assessment: The pilot year.* Montpelier, VT: Author.

Vermont Department of Education. (1991a). *"This is my best": Vermont's Writing Assessment Program, pilot year 1990-1991.* Montpelier, VT: Author.

Vermont Department of Education. (1991b). *Looking beyond "The Answer": The report of Vermont's Mathematics Portfolio Assessment Program.* Montpelier, VT: Author.

Vermont Department of Education. (1991c). *Vermont Mathematics Portfolio Project: Resource book.* Montpelier, VT: Author.

Vermont Department of Education. (1991d). *Vermont Mathematics Portfolio Project: Teacher's guide.* Montpelier, VT: Author.

Webb, N. (1993). *Collaborative group versus individual assessment in mathematics: Group processes and outcomes* (CSE Tech. Rep. 352). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Webb, N. M. (1994). *Group collaboration in assessment: Competing objectives, processes, and outcomes* (CSE Tech. Rep. 386). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Wolf, D. P. (1989, April). Portfolio assessment: Sampling student work. *Educational Leadership, 46*(7), 4-10.

Portfolio Assessment: *Whose Work Is It?*
Issues in the Use of Classroom Assignments for Accountability

Wolf, D. P. (1993). Assessment as an episode of learning. In R. Bennett & W. Ward (Eds.), *Construction vs. choice in cognitive measurement* (pp. 213-240). Hillsdale, NJ: Lawrence Erlbaum Associates.

Wolf, D. P., Bixby, J., Glenn, J., & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. In G. Grant (Ed.), *Review of Research in Education* (Vol. 17, pp. 31-74). Washington, DC: American Educational Research Association.

Wolf, S. A., & Gearhart, M. (1993a). *Writing what you read: Assessment as a learning event* (CSE Tech. Rep. 358). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Wolf, S. A., & Gearhart, M. (1993b). *Writing what you read. A guidebook for the assessment of children's narratives* (CSE Resource Paper No. 10). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Yancey, K. B. (1992). Portfolios in the writing classroom: A final reflection. In K. B. Yancey (Ed.), *Portfolios in the writing classroom: An introduction* (pp. 102-116). Urbana, IL: NCTE.

[1] Both authors contributed equally to this article. Portions of this article are closely adapted from Gearhart, Herman, Baker, & Whittaker (1993). We thank Noreen Webb, Dan Koretz, Brian Stecher, Drew Gitomer, and Ron Dietel for detailed comments and revisions of an earlier draft.

[2] Examples include: Baker, Gearhart, Herman, Tierney, & Whittaker, 1991; Belanoff & Dickson, 1991; Calfee & Perfumo, 1992, 1993a, 1993b; Camp, 1990, 1993; Camp & Levine, 1991; Freedman, 1993; Gentile, 1992; Gill, 1993; Gitomer, 1993; Glazer, Brown, Fantauzzo, Nugent, & Searfoss, 1993; Hamp-Lyons & Condon, 1993; Herman, Aschbacher, & Winters, 1992; Herman, Gearhart, & Aschbacher, in press; Hewitt, 1991, 1993; Hiebert & Calfee, 1992; Hill, 1992; Howard, 1990; Koretz, 1993; Koretz, Klein, McCaffrey, & Stecher, 1993; Koretz, McCaffrey, Klein, Bell, & Stecher, 1993; Koretz, Stecher, & Deibert, 1992; Koretz, Stecher, Klein, & McCaffrey, in press; Koretz, Stecher, Klein, McCaffrey, & Deibert, 1993; LeMahieu, Eresh, & Wallace, 1992; LeMahieu, Gitomer, & Eresh, in press; Meyer, Schuman, & Angello, 1990; Mills, 1989; Mills & Brewer, 1988; Moss, 1992; Moss, Beck, Ebbs, Herter, Matson, Muchmore, Steele, & Taylor, 1991; Murphy & Smith, 1990, 1992; O'Neil, 1992; Paulson, Paulson, & Meyer, 1991; Reidy, 1992; Saylor & Overton, 1993; Sheingold 1994; Simmons, 1990; Smith & Murphy, 1992; Stecher & Hamilton, 1994; Stroble, 1993; Tierney, 1992; Tierney, Carter, & Desai, 1991; Valencia & Calfee, 1991; Vermont Department of Education, 1990, 1991a, 1991b, 1991c, 1991d; Wolf, D. P., 1989, 1993; Wolf, Bixby, Glenn, & Gardner, 1991; Yancey, 1992.

[3] Moss (1994) has proposed that high-stakes, large-scale writing assessment could be conducted in ways that afford greater local authority and that foster dialogue and ongoing review of both the methods of portfolio assessment and the results. Her work is the impetus for very productive debate concerning a complex set of issues, including the ways that the goals and models of portfolio assessment programs may need to be designed differently for different levels of an assessment system. In this paper, we reify the features of a "large-scale portfolio assessment program," both simplifying their complexity and drawing on features of several current programs.

[4] In the Kentucky program, the classroom teacher scores his or her own portfolios, but a sampling of portfolios is rescored by another teacher (usually from another school) and/or by a KRIS staff member. Because a given teacher's scores may be adjusted by patterns of relationships between his scores and an outsider's, we still view the issue for Kentucky in the same way: Credibility for a student's score still derives ultimately from an outsider's judgment.

[5] Koretz, 1993; Koretz, Klein, McCaffrey, & Stecher, 1993; Koretz, McCaffrey, Klein, Bell, & Stecher, 1993; Koretz, Stecher, & Deibert, 1992; Koretz, Stecher, Klein, & McCaffrey, 1994, in press; Koretz, Stecher, Klein, McCaffrey, & Deibert, 1993.

[6] There are certainly other possible explanations. Moss (1994), for example, discusses potential incompatibilities among methods and purposes of different approaches to assessment.

[7] Again we are assuming a rater who is scoring outside the classroom context.

[8] The description below is taken from Koretz et al., 1994. Readers are referred to that technical report for more details about the study.

[9] The description below is taken from Gearhart et al., 1993. Readers are referred to that technical report for more details about the study.

[10] To compensate for variation in the number of assignments rated per child and the number of children designated as high, medium, or low in writing ability, weighted averages were computed for each teacher. Thus, a teacher's ratings were averaged for each student's assignments, and then, as appropriate, a "mean of means" was computed for each teacher, or for "high," "medium," and "low" students for each teacher.

New & Recent CRESST/CSE Technical Reports

On Concept Maps as Potential "Authentic" Assessments in Science

Richard Shavelson, Heather Lang, and Bridget Lewin
CSE Technical Report 388, 1994 ($4.00)

Using concept maps as an assessment tool will urge educators to teach students more than simple facts and concepts, but how different concepts relate to each other. An evaluational tool such as concept mapping urges the individual to think on a deeper cognitive level than a "fill-in the blank" test would require. There is value in both assessment tools—neither should be ignored.

Fourth-grade teacher

The preceding statement from *On Concept Maps as Potential "Authentic" Assessments in Science* points to the multiple expected benefits from the use of concept maps as alternative assessments. In this report, CRESST researchers Richard Shavelson, Heather Lang, and Bridget Lewin examine concept mapping issues as an assessment technique exploring questions related to validity and reliability of student scores. The authors begin with a clear definition of concept mapping:

"A concept map," write the authors, "constructed by a student, is a graph consisting of nodes representing concepts and labeled lines denoting the relation between a pair of nodes (concepts)."

Based on an extensive review of concept map usage, the authors found that concept mapping techniques differed widely.

"No less than 128 possible variations were identified" say the authors. "Methods for scoring maps varied almost as widely, from the admonition 'don't score maps' to a detailed scoring system for hierarchical maps."

The researchers' review led to the conclusion that an integrative "working" cognitive theory is needed to begin to limit this great variation for alternative assessment purposes. "Such a theory," conclude the authors, "would also serve as a basis for much needed psychometric studies of the reliability and construct validity of concept maps since such studies are almost nonexistent in the literature."

The review presents issues arising from large-scale use of mapping techniques, including the importance of students' skills in using concept maps, and the possible negative impact of teachers teaching to the assessment if students have to *memorize* concept maps provided by textbooks or themselves.

Specifications for the Design of Problem-Solving Assessments in Science

Brenda Sugrue
CSE Technical Report 387, 1994 ($4.00)

In *Specifications for the Design of Problem-Solving Assessments in Science*, CRESST researcher Brenda Sugrue draws on the CRESST performance assessment model to develop a new set of test specifications for science. Sugrue recommends that designers follow a straightforward approach for developing alternative science assessments.

"Carry out an analysis of the subject matter content to be assessed," says Sugrue, "identifying key concepts, principles, and procedures that are embodied in the content." She adds that much of this analysis already exists in state frameworks or in the national science standards.

Either multiple-choice, open-ended, or hands-on science tasks can then be created or adapted to measure individual constructs, such as concepts and principles, and the links between concepts and principles.

In addition to measuring content-related constructs, Sugrue's model advocates measuring metacognitive constructs and motivational constructs in the context of the content. This permits more specific identification of the sources of students' poor performance. Students may perform poorly because of deficiencies in content knowledge, and/or deficiencies in constructs such as planning and monitoring, and/or maladaptive perceptions of self

and task. The more specific the diagnosis of the source of poor performance, the more specific can be instructional interventions to improve performance.

Sugrue's model includes specifications for task design, task development, and task scoring, all linked to specific components of problem-solving ability. An upcoming CRESST report will discuss the results of a study designed to evaluate the effectiveness of the model for attributing variance in performance to particular components of problem solving and particular formats for measuring them.

Group Collaboration in Assessment: Competing Objectives, Processes, and Outcomes
Noreen Webb
CSE Technical Report 386, 1994 ($4.00)

Learning from other students, developing interpersonal skills, and maximizing collaborative performance are three primary goals of small-group collaboration. But according to Noreen Webb in *Group Collaboration in Assessment: Competing Objectives, Processes, and Outcomes*, group assessment may or may not be an effective strategy for measuring such goals.

"There may be an appropriate place for collaborative group work in educational assessment," asserts Webb. "Most importantly," she adds, "how group work is used in an assessment should coincide with the purpose of an assessment."

Webb's own research and that of others strongly suggests that the purposes must be clearly understood from the start of a group assessment project. Measuring individual student learning versus group productivity, for example, may call for differences in the assessments.

If the purpose is to measure individual achievement, suggests Webb, then the instructions might be worded to encourage individual effort. A Connecticut assessment, for example, told students that "each person should be able to explain fully the conclusions reached by the group" and should be prepared to give an oral presentation on their group's experiment. Thus, students expected that they would be held accountable individually.

But if the assessment purpose is different, so too should be the focus, says Webb. She cites group productivity as one example:

"Assessment focusing on group productivity," says Webb, "would give a group a task to complete, and evaluation would focus on the completed task, not on individual students' contributions to completing the task."

Webb says that teachers need to prepare students for group assessment. Small-group collaboration can help students to develop valuable communications skills and give them a better "understanding of what kinds of group processes help them learn as individuals and what kinds of processes help maximize group productivity," she concludes.

The Evolution of a Portfolio Program: The Impact and Quality of the Vermont Program in Its Second Year
Daniel Koretz, Brian Stecher, Stephen Klein, and Daniel McCaffrey
CSE Technical Report 385, 1994 ($4.00)

Part of an ongoing evaluation of the Vermont portfolio assessment program by RAND/CRESST researchers, this report presents recent analyses of the reliability of Vermont portfolio scores, and the results of school principal interviews and teacher questionnaires.

The message, especially from Vermont teachers, say the researchers, remains mixed. Math teachers, for example, have modified their curricula and teaching practices to emphasize problem solving and mathematical communication skills, but many feel they are doing so at the expense of other areas of the curriculum. About one-half of the teachers report that student learning has improved, but an equal number feel that there has been no change. Addi-

tionally, teachers reported great variation in the implementation of portfolios into their classrooms, including the amount of assistance provided to students.

"One in four teachers," found the authors, "does not assist his or her own students in revisions, and a similar proportion does not permit students to help each other. Seventy percent of fourth-grade teachers and 39% of eighth-grade teachers forbid parental or other outside assistance."

Consequently, students who receive more support from teachers, parents and other students may have a significant advantage over students who receive little or no outside help.

Reliability problems continue. "The degree of agreement," write the authors, "among Vermont's portfolio raters was much lower than among raters in studies with other types of constructed response measures."

The authors suggest that one cause of the low reliability was the diversity of tasks within each portfolio. Because teachers and students are free to select their own pieces, performance on the tasks is much more difficult to assess than if the work were standardized.

Despite these problem areas, support for the portfolio program remains high. Teachers, for example, expressed strong support for expanding portfolios to all grade levels. Seventy percent of principals said that their schools had extended portfolio usage beyond the original Vermont state mandate.

A Conceptual Framework for Analyzing the Costs of Alternative Assessment

Lawrence O. Picus

CSE Technical Report 384, 1994 ($4.00)

Despite the fact that many states are investing millions of dollars in the development of alternative assessments, little is known about the actual costs of such assessments. In *A Conceptual Framework for Analyzing the Costs of Alternative Assessments*, CRESST partner Lawrence O. Picus analyzes many of the issues related to identifying the costs of new assessments including the relationship between costs and goals.

"If, as is often the case in education," says Picus, "there are multiple goals established for an alternative assessment program, then estimation of the costs of that program must include all of the resources necessary to accomplish all of those goals."

To identify alternative assessment costs, Picus suggests the use of a three-dimensional model comprised of *levels* of expenditures, *kinds* of expenditures, and expenditure *components*. Levels of expenditures are the source of expense such as national, state, district, school, classroom, or private market levels. Kinds of expenditures include personnel, materials, supplies, and travel. Components include assessment development, production, training, scoring, reporting and program evaluation.

"The largest single expenditure item in any assessment program," concludes Picus, "seems likely to be personnel."

Opportunity costs must also be considered, adds Picus. Resources committed to creating an alternative assessment program are resources used to support a former testing program or resources that could be spent on other programs, such as bilingual education.

The framework developed in *A Conceptual Framework for Analyzing the Costs of Alternative Assessment* addresses both opportunity costs and assessment costs matched to goals.

Recent CRESST/CSE Technical Reports

Economic Analysis of Testing: Competency, Certification, and "Authentic" Assessments
James S. Catterall and Lynn Winters
CSE Technical Report 383, 1994 ($4.00)

Cost-benefit and cost-effectiveness analyses, say researchers James Catterall and Lynn Winters in *Economic Analysis of Testing: Competency, Certification, and "Authentic" Assessments*, have similar policy purposes.

"Both analyses," note the authors, "aim at what choices might be made either to reach given goals with lower costs, or to attain more results for a given budget allocation."

But trying to use either economic analysis is difficult when applied to educational assessment because anticipated benefits are moot. The authors note, for example, that policy makers, test coordinators, principals, and counselors have used minimum competency tests to motivate students, encouraging them, albeit negatively, to develop and improve basic skills. Yet a study by James Catterall in 1990 showed that fewer than half of 736 students in eight high schools (four states) were even aware that their high school required them to pass a minimum competency test prior to graduation. Clearly the intended motivational benefit or effect sought by policy makers was not reflected by what was truly happening. Thus, tying the costs of assessments to benefits and effects that may not really occur is problematic.

Regarding performance assessments, the authors suggest that linking policy and costs will be equally challenging. Because performance assessments should be good instructional activities themselves, it is difficult to differentiate the costs into specific categories such as assessment, curriculum, or, in cases of scoring, professional development for teachers.

"That they [performance assessments] have returns over and above current tests is presently assumed," conclude Catterall and Winters. "Establishing the linkages between the costs and benefits may be an important factor in the course of testing reform in the 1990s."

Analysis of Cognitive Demand in Selected Alternative Science Assessments
Gail Baxter, Robert Glaser, and Kalyani Raghavan
CSE Technical Report 382, 1994 ($4.00)

Working with pilot science assessments in California and Connecticut, the researchers in *Analysis of Cognitive Demand in Selected Alternative Science Assessments* focused on cognitive activity required for successful completion of performance assessment tasks. Of special interest was the degree to which task performance reflected differences in student understanding.

"We focused," wrote Baxter, Glaser, and Raghavan, "on the extent to which: (a) tasks allowed students the opportunity to engage in higher order thinking skills and (b) scoring systems reflected differential performance of students with respect to the nature of cognitive activity in which they engaged."

Data came from three types of science assessment tasks–exploratory investigation, conceptual integration, and component identification–each varying with respect to grade level, prior knowledge, stage of development and purpose. Analyses of the data resulted in some important recommendations for the development of assessment tasks and scoring.

In general, wrote the authors, "tasks should: (a) be procedurally open-ended affording students an opportunity to display their understanding; (b) draw on subject matter knowledge as opposed to knowledge of generally familiar facts; and (c) be cognitively rich enough to require thinking."

The authors concluded that scoring systems should: (a) link score criteria to task expectations; (b) be sensitive to the meaningful use of knowledge; and (c) capture the problem-solving processes the students engage in.

Recent CRESST/CSE Technical Reports

Measurement-Driven Reform: Research on Policy, Practice, Repercussion

Audrey J. Noble and Mary Lee Smith

CSE Technical Report 381, 1994 ($4.00)

Demonstrating a top-down educational reform strategy and a belief that assessment can leverage educational change, the Arizona state legislature in 1990 passed Arizona Revised Statute 15-741. The legislation resulted in the Arizona Student Assessment Program (ASAP), a program that incorporated both standardized and performance-based assessments. *Measurement-Driven Reform: Research on Policy, Practice, Repercussion* reports on how ASAP was conceived, negotiated, and implemented. The CRESST researchers conducting the study, Audrey Noble and Mary Lee Smith, were critical of the policy process that created ASAP.

ASAP "reveals both the ambiguities characteristic of the [assessment] policy-making process," write Noble and Smith, "and the dysfunctional side effects that evolve from the policy's disparities."

Employing multiple research methods, the researchers interviewed members of the policy-shaping community and examined documents and artifacts of the testing policy. Their analysis determined that competing ideas about student learning, teachers, curriculum, and assessment resulted in ineffective implementation of the assessment program. Five inconsistencies were reported:

- Policy makers' definitions of "learning" were incoherent;
- Policy makers held dissonant expectations of teachers;
- Policy makers clashed regarding the role of curriculum;
- Policy makers alleged that a single performance assessment could fulfill the dual purposes of instructional improvement and accountability;
- The implementation plan of the Arizona Student Assessment Program was a dysfunctional side effect of a policy built on contradictory ideals.

"Although ASAP appeals to many because of its ambiguity," conclude Noble and Smith, "this same characteristic may undermine its capacity to effect any substantial change in educational practice."

What Happens When the Test Mandate Changes? Results of a Multiple Case Study

Mary Lee Smith, Audrey J. Noble, Marilyn Cabay, Walt Heinecke, M. Susan Junker, and Yvonne Saffron

CSE Technical Report 380, 1994 ($4.50)

The Arizona Student Assessment Program (ASAP) was designed to solve what policy makers perceived to be the state's most pressing educational problems: moving schools toward the state curriculum framework and making schools more accountable for student achievement. However, as findings from this multiple case study demonstrate, the actions of practitioners were far from uniform in response to this policy mandate.

In this report, CRESST researcher Mary Lee Smith and colleagues outline the results to date of a three-year, qualitative study of school reactions to the ASAP mandate. One of a series of reports on a larger project, *What Happens When the Test Mandate Changes?*, the present study addresses the consequences of the change mandate in four Arizona elementary schools during the first year of implementation.

Using a case study methodology, the researchers focused on the interplay of policy and practice by engaging directly in the local, school-site scene. This particular approach allowed them to gain access to participant meanings and to show how meanings-in-action evolved over time.

New & Recent CRESST/CSE Technical Reports

Results from this study indicated that local school responses to the policy mandate varied substantially. The goal of transforming classroom instructional practices was achieved in only one school that had adopted such practices prior to the state mandate.

"Local interpretations and organizational norms intervened to color, distort, delay, enhance, or thwart the intentions of the policy and the policy-shaping community," concluded the authors. Expectations for school reform based on mandates must consider the vast disparities that exist between individual schools and teachers.

Assessment, Testing, and Instruction: Retrospect and Prospect
Robert Glaser and Edward Silver
CSE Technical Report 379, 1994 ($4.50)

Increasing concern about the nature and form of student assessments and the uses made of test results forms the basis for *Assessment, Testing and Instruction: Retrospect and Prospect* by CRESST researchers Robert Glaser and Edward Silver. The authors explore the nature of testing and assessment by examining some of the deficiencies and abuses associated with past practices in educational measurement, then investigating present and future possibilities for alternative forms of assessment.

"At this point in time," write Glaser and Silver, "assessment and testing in American schools are caught between the extensive rhetoric of reform and the intransigence of long-established practices." Through an informative discussion of the two standard purposes of educational assessment—testing for selection and placement and assessing educational outcomes—the authors demonstrate the need for an evaluation of the purposes of educational testing.

MORE TECHNICAL REPORTS

Policy Makers' Views of Student Assessment
Lorraine McDonnell
CSE Technical Report 378, 1994 ($4.00)

Engaging Teachers in Assessment of Their Students' Narrative Writing: Impact on Teachers' Knowledge and Practice
Maryl Gearhart, Shelby A. Wolf, Bette Burkey, and Andrea K. Whittaker
CSE Technical Report 377, 1994 ($4.00)

Test Theory Reconceived
Robert J. Mislevy
CSE Technical Report 376, 1994 ($4.50)

Linking Statewide Tests to the National Assessment of Educational Progress: Stability of Results
Robert L. Linn and Vonda L. Kiplinger
CSE Technical Report 375, 1994 ($4.00)

UCLA's Center for the Study of Evaluation & The National Center for Research on Evaluation, Standards, and Student Testing
Eva L. Baker, Co-director
Robert L. Linn, Co-director
Joan L. Herman, Associate Director
Ronald Dietel, Editor
Katharine Fry, Editorial Assistant

The work reported in this publication was supported under the Educational Research and Development Center Program cooperative agreement number R117G10027 and CFDA catalog number 84.117G as administered by the Office of Educational Research and Improvement, U.S. Department of Education. The findings and opinions expressed in this publication do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

OTHER CURRENT CRESST/CSE PRODUCTS

DATABASES

Alternative Assessments in Practice Database
Listings from over 250 developers of new assessments, Macintosh version only, 1993 ($15.00)

VIDEOTAPES

Assessing the Whole Child
18-minute video program includes new practitioner's guidebook, V3, 1994 ($15.00)

Portfolio Assessment and High Technology
10 minute video program includes practitioner's guidebook on portfolio assessment, V2, 1992 ($12.50)

HANDBOOK

CRESST Performance Assessment Models: Assessing Content Area Explanations
Eva L. Baker, Pamela R. Aschbacher, David Niemi, and Edynn Sato
The CRESST Handbook, 1992 ($10.00)

RESOURCE PAPER

Writing What You Read: A Guidebook for the Assessment of Children's Narratives
Shelby Wolf and Maryl Gearhart
CSE Resource Paper 10, 1993 ($4.00)

For a complete list of all CRESST products, please contact Kim Hurst at 310-206-1532, e-mail: kim@cse.ucla.edu, or UCLA Center for the Study of Evaluation, 10880 Wilshire Blvd., #700, Los Angeles, CA 90024-1394.

FOLD AND SECURE

Place Postage Here

UCLA Center for the Study of Evaluation
10880 Wilshire Blvd., Suite 700
Los Angeles, California 90024-1394

Order Form

Attach additional sheet if more room is needed. Form is pre addressed on reverse

CSE/CRESST Products

CSE Number	Title	Number of Copies	Price per Copy	Total Price

POSTAGE & HANDLING
(Special 4th Class Book Rate)

Subtotal of	$0 to $10	add $1.50
	$10 to $20	add $2.50
	$20 to $50	add $3.50
	over $50	add 10% of Subtotal

ORDER SUBTOTAL _____

POSTAGE & HANDLING (scale at left) _____

California residents add 8.25% _____

TOTAL _____

Orders of less than $10.00 must be prepaid

Your name & mailing address—please print or type:

_____

_____

_____

❑ Payment enclosed ❑ Please bill me

❑ I would like to receive free copies of the *CRESST Line* and *Evaluation Comment* publications.

UCLA Center for the Study of Evaluation
10880 Wilshire Blvd., Suite 700
Los Angeles, California 90024-1394

ADDRESS CORRECTION REQUESTED (ED 63)

NONPROFIT ORG.
U.S. POSTAGE
PAID
U.C.L.A.

Dr. Carol Ascher
Senior Research Assoc.
Institute for Urban & Minority Ed.
Columbia University Teachers College
P.O. Box 40
New York NY 10027